

Weekly Report

April 14, 2019

1 Work

1. 本周同时在进行unpair setting下的图片增强，和基于空间位移的Adversarial Attack，目前还在测试思路的可能性。
2. DRGraph已经完成了其他算法的测试代码。
3. 工作时长：工作日每天10个小时，周末共10个小时，共60个小时。

1.1 工作进度

Table 1: 工作进度

项目	进度	截止时间
DRGraph	需要对程序做一些修改	2019.4.30
unpair 低光照图片增强	目前初步的实验效果不佳	
NIPS	Adversarial Attack	2019.5.23

2 Paper Reading

2.1 Extracting Relationships by Multi-Domain Matching

有时候收集到的数据往往存在着一些bias，比如病人数据，将其运用到新的病人身上时就会没有那么适用。所以文章提出Multi-Domain Matching，根据多个数据集之间的相似性来拉近不同数据集在特征空间中的距离，然后在此基础上学习一个分类器。

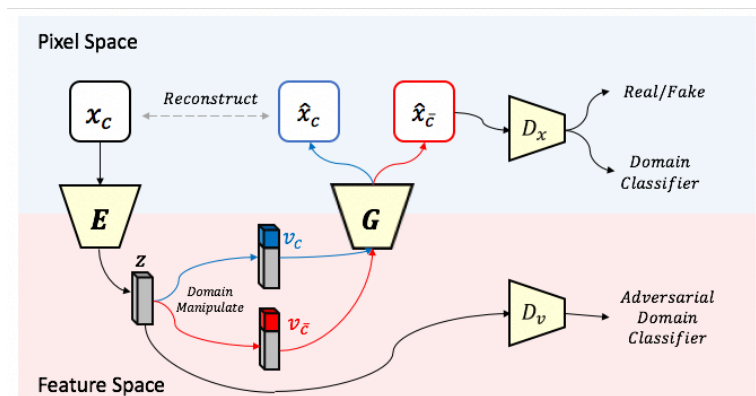


Figure 1: #1

2.2 Text-Adaptive Generative Adversarial Networks: Manipulating Images with Natural Language

基于自然语言修改图片，主要是提出了text-adaptive discriminator 基于word层次的判别网络。

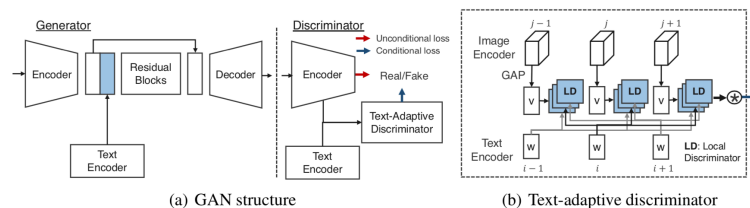


Figure 2: #2

2.3 Universal adversarial perturbations

计算了一个适用于整个数据集的对抗扰动，而不是对每一个图片单独计算。

2.4 SPATIALLY TRANSFORMED ADVERSARIAL EXAMPLES

通过位移像素点而不是加噪声的方法也可以达到生成对抗样本的同时保持图片尽可能真实。

2.5 Explaining and Harnessing Adversarial Examples (ICLR2015)

如果输入数据受到对抗扰动影响，即 $x+e$ ，则在现行模型的结果就会出现 $W(x+e)=Wx+W_e$ ， W_e 则是对抗扰动对结果的影响，其中 e 取 W 的符号。把类似的方法用到神经网络中，也同样适用，其中 W 可以看成是神经网络对于 W 的梯度，等价于increase the cost of the correct class。

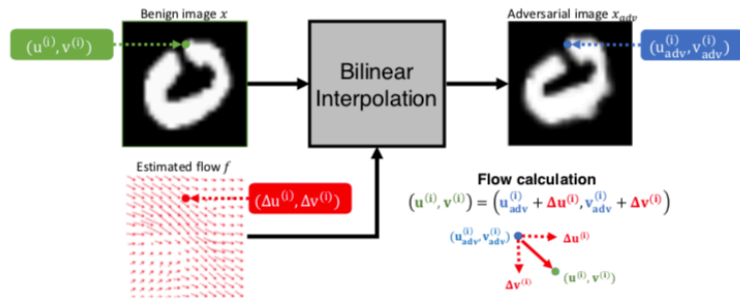


Figure 3: #4